Riphah Journal of Islamic Thought & Civilization
Published by: Department of Islamic Studies,
Riphah International University, Islamabad
Email: editor.rjitc@riphah.edu.pk
Website: https://journals.riphah.edu.pk/index.php/jitc
ISSN (E): 3006-9041 (P): 2791-187X

Tabayyun and Qurʾānic Digital Ethics against Misinformation and Hate Speech in Pakistan

Jamil Akhtar[1]

Abstract:

This paper discusses how the Qurʾānic principles of tabayyun (responsible verification) can be formulated as a digital ethics framework to address fake news and hate speech in the Pakistani online community. The research will (i) develop an informative Qurʾānic conceptual model of information ethics, (ii) evaluate how prevalent trends of online misinformation and hate speech contravene Qurʾānic norms of communication, and (iii) present ethically relevant suggestions to individuals, religious leadership, and policy makers. The article methodologically uses a thematic Qurʾānic analysis of verses on verification, truthful speech, slander, suspicion and social harmony, with a qualitative content analysis of a few cases of publicly available misinformation and hate speech, originating in Pakistan. The results suggest that the Qurʾān encodes information as a moral trust with a sense of responsibility, forbids reputational damage and identity-based contempt, and requires verification, particularly when a person is most likely to harm others. The research wraps up by recommending a workable "tabayyun protocol" for digital sharing and defining institutional actions to media literacy, khutbah discourse and responsible regulation.

Keywords: *Digital misinformation; Hate speech; Islamic communication ethics; Qurʾānic ethics; Tabayyun*.

*Assistant Professor, Urdu Encyclopedia of Islam, University of the Punjab, Lahore, Pakistan.
Email ID: jamillahori@gmail.com
ORCID: https://orcid.org/0000-0001-9884-4348
DOI: https://doi.org/10.64768/rjitc.v4i1.2955

1. Introduction:

The digital sphere in Pakistan is now one of the main arenas where meanings of religion, sectarianism, and politics are created and disputed. However, it is the same networked environment that facilitates the rapid distribution of valuable information as well as misinformation and hate speech, and in many cases, it comes with high-speed sharing, anonymity, and emotional engagement. Modern academia characterizes this ecosystem as "information disorder," in which fake or misleading information is propagated not only through ignorance but also with the encouragement of platform engineering and social polarization, where false allegations on the internet may turn into real-life confrontation.[1] Simultaneously, hate speech is labeled by the international policy discourse as a source of discrimination and physical harm in the real world, in particular, when it entails mobilizing identity contempt and incitement.[2] The dynamics from which they are often conflated in Pakistan include religious sensitivities, reputation vulnerability, and flimsy communal relationships, which allow unverified accusations to develop into offline conflicts of the online accusation.

This paper suggests that the Qurʾānic doctrine of *tabayyun* (تبيّن)—ethical verification preceding acceptance, transmission, or action—can provide more than just ethical advice; it can be considered to offer a consistent ethics of digital life when interpreted in connection with other relevant Qurʾānic norms of speech, dignity, and justice. Verification is made an obligation in Q 49/6, when the potentially damaging *naba'* (news) is present, as it has always been stressed in the classical tafsīr literature.[3] Communicative integrity is also grounded in knowledge-claim accountability (Al-Qurʾān: 17/36), disciplined truth-telling (*qawlsadīd*, Q 33/70), and harsh reproach toward mockery, suspicion, and indemnity (Al-Qurʾān: 49/11-12) in the Qurʾān. The moral template of the episode of *al-ifk*(Al-Qurʾān: 24/11-19) is especially relevant to

[1] Claire Wardle and Hossein Derakhshan, *Information Disorder Toward an Interdisciplinary Framework for Research and Policy Making* (Strasbourg: Council of Europe, 2017), p. 39. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.

[2] Office of the United Nations High Commissioner for Human Rights (OHCHR), *Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred that Constitutes Incitement to Discrimination, Hostility or Violence* (Geneva: OHCHR, 2012), https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf.

[3] Muḥammad ibn Jarīr al-Ṭabarī, *Jāmiʿ al-bayānʿantaʾwīlāy al-Qurʾān* (تفسير الطبري: جامع البيان عن تأويل آي القرآن), commentary on Q 49:6.

digital culture, where widely disseminated alleged information legitimizes public shaming and postpones its rectification.[1]

The research questions of this paper are threefold: (i) to develop a Qurʾān-based conceptual framework of digital information ethics based on the principle of *tabayyun* and its close allies; (ii) to evaluate how widely spread misinformation and hate speech in Pakistan contravene Qurʾānic standards of communication; and (iii) to make practical conclusions about what individuals, teachers, religious figures, and institutions should do.

Thematic Qurʾānic analysis (*tafsīr mawḍūʿī*), consisting of semantic consideration of key ethical terms (e.g., *tabayyun*, *ʿilm*, *buhtān*, *ʿadl*, *iṣlāḥ*) as the methodological approach, is supplemented with a qualitative content analysis of examples of misinformation and hate-speech focusing on Pakistan, recorded by reputable fact-checking processes, and this allows them to focus on a contemporary challenge with analytical clarity and applied relevance.[2] The unique contribution relates to the concept of turning Qurʾānic moral vocabulary into a working "*tabayyun* protocol" of the digital sharing of materials so that Qurʾānic studies can respond to a modern challenge in an analytically consistent and practically relevant way.

2. Research Framework and Conceptual Model:

The current paper adheres to the interdisciplinary normative-analytical framework of study, in which the instructions of the Qurʾān are interpreted not only as a devotional guide but as a coherent moral epistemology to be applied in solving secular problems externally. The starting point is the contemporary definition of information disorder, in which fake, misleading, or weaponized content is sent and received by social incentives (identity performance, outrage) and platform incentives (visibility, algorithmic amplification).[3] In this context, there is an analytical distinction between "misinformation" (true content published without the intent to cause harm) and "disinformation" (true content published with the intent of causing harm) that is ethically convergent when circulation can be predictably followed with causing public

[1] Ismāʿīl ibn Kathīr, *Tafsīr al-Qurʾān al-ʿAẓīm* (تفسير القرآن العظيم: تفسير ابن كثير), commentary on Q 24:11–19.
[2] Soch Fact Check, "Soch Fact Check," accessed February 14, 2026, https://www.sochfactcheck.com/; Agence France-Presse (AFP), "AFP Fact Check Pakistan," accessed February 14, 2026, https://factcheck.afp.com/afp-pakistan.
[3] Wardle and Derakhshan, *Information Disorder*, p. 76.

harm.[1] In the case of hate speech, the study is based on internationally accepted definitional literature, which focuses on preaching hostility and incitement toward guarded identities, although applicable legal standards differ.[2] The Qurʾānic intervention here is to consider digital communication as mukallaf (ethically responsible) action, the ethicality of which revolves around the diligence of epistemology (*tabayyun*), the honesty of speech (*qawl* ethics), and prevention of harm (avoiding *fasād*).

2.1 Key Qurʾānic concepts as analytical categories:

The theoretical framework is based on thematic Qurʾānic reading (*tafsīrmawḍūʿī*) and semantic interest in terms regulating the ethics of accepting and presenting claims. Q 49/6 refers to the anchor as *tabayyun* (تبيّن), which is a duty of exegetes of that era to determine the status of a report (*nabaʾ*) before proceeding to a course of action that could harm others.[3] Concurrently, Q 17/36 denies the seeking of claims without knowledge (*ʿilm*) and makes hearing, sight, and heart explicitly liable. This verse is interpreted in the model as a broad epistemic limitation in the context of "share-first" digital culture. The discourse ethics of Qurʾānic speech also demand the presence of *qawlsadīd* (قول سديد) in Q 33/70: a speech that is precise, disciplined, and socially responsible, while Q 24/15-16 condemns the casually circulating rumors and the implements of moral insignificance of unverified speech.[4]

The third conceptual axis is protection of human dignity (*ʿirḍ*, عِرض) and privacy, derived from Q 49/11-12 banning mockery (*sukhriyyah*), name-calling (*tanābuzbil-alqāb*), suspicion (*ẓann*), and backbiting (*ghībah*). The passage of the *ifk*(Al-Qurʾān: 24/11-19) is here a sort of Qurʾānic "case study" on the issue of reputational harm and corrective responsibility: the spread of accusations among society is considered a social moral failure, rather than an individual wrong.[5] The

---

[1] Cherilyn Ireton and Julie Posetti, eds., *Journalism, "Fake News" & Disinformation Handbook for Journalism Education and Training* (Paris: UNESCO, 2018), https://unesdoc.unesco.org/ark:/48223/pf0000265552.

[2] Office of the United Nations High Commissioner for Human Rights (OHCHR), *Rabat Plan of Action*; United Nations, *United Nations Strategy and Plan of Action on Hate Speech* (New York: United Nations, 2019), https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf.

[3] Abūʿ Abd Allāh Muḥammad ibn Aḥmad al-Qurṭubī, *al-Jāmiʿ li-aḥkām al-Qurʾān* (تفسير القرطبي: الجامع لأحكام القرآن), commentary on Q 49:6; Fakhr al-Dīn al-Rāzī, *Mafātīḥ al-ghayb* (تفسير الرازي: مفاتيح الغيب), commentary on Q 49:6.

[4] Ibn Kathīr, *Tafsīr al-Qurʾān al-ʿAẓīm*, commentary on Q 24:15–16.

[5] al-Ṭabarī, *Jāmiʿ al-bayān*, commentary on Q 24:11–19.

fourth is justice (*ʿadl*) in Q 5/8, directly opposing group antagonism as a justification to unfair judgment—very applicable to sectarian or partisan "selective verification." Another fifth axis is the social repair (*iṣlāḥ*) in Q 49/10 that defines communicative responsibility as the de-escalation and cohesion rather than mere victory in a dispute.

2.2 The Qurʾānic Digital Ethics Model of *Tabayyun* (QDEMT):

Through these conceptual axes, the article proposes the Qurʾānic Digital Ethics Model of *Tabayyun* (QDEMT) to be an information lifecycle that has three ethically handled phases:

a. Receive (Reception Ethics): The digital exposure is deemed morally pertinent as it invokes judgment. QDEMT assumes (a) that the claim (news, accusation, religious warning, call to action) has to be categorized and (b) the categories of "high-harm" content (accusations, identity-targeting content, mobilization posts) must be more stringently verified.

b. Verify (Epistemic Due Diligence): This has operationalized *tabayyun* to mean source auditing (by whom did they claim to say it and with what traceability), evidence auditing (what shall pass as proof, situation, and cooperation), and harm auditing (who will be harmed by publication). This step represents Qurʾānic epistemic interdictives (Al-Qurʾān: 49/6; Q 17/36) in the form of observable and qualifiable behavior that can be used in qualitative codification (e.g., existence of primary evidence, corroboration, contextualization, necessity of not being basically true without evidence).

c. Transmit or Respond (Speech Integrity and Harm Prevention): In case of failure to verify or probable harm, restraint is an ethical obligation. The acceptable correction should be based on *qawlsadid* and dignity principles—correcting the misinformation but should not use the language of contempt and should avoid public admonishment where legal means are available (Al-Qurʾān: 24/11-19; Q 49/11-12).

Table 1: QDEMT Conceptual Framework Matrix for *Tabayyun* and Qurʾānic Digital Ethics

QDEMT stage	Core Qurʾānic anchor	Ethical objective	Required ethical actions in digital	High-risk Pakistan digital patterns	Qurʾān-aligned decision rule for	Observable indicators for

	s		practice	addressed	the user	qualitative analysis
Stage 1 Receive (Reception ethics)	Q 49/6; Q 17/36	Prevent harm caused by impulsive belief and moral panic	Pause before reacting; classify the claim type (news, accusation, mobilization, "religious warning"); recognize "high-harm" content as requiring stricter verification	"Forward urgently"; screenshot-driven claims; clipped videos; emotionally charged captions; sectarian bait framing	If the claim is accusatory or mobilizing and the source is unclear, do not share and do not judge until verification is complete	Urgency cues; emotional manipulation; anonymous forwarding; lack of original source; immediate calls to act
Stage 2 Verify (*Tabayyun* and epistemic due diligence)	Q 49/6; Q 17/36; Q 33/70	Establish truth-responsibility and evidence-based judgment	Source audit (who originates, traceability, credibility); evidence audit (primary proof, corroboration, context); harm audit (dignity risk, sectarian escalation, public safety)	Virality replacing verification; piety-coded claims treated as unquestionable; "selective verification" for in-group narratives	Verification must be proportional to harm potential the greater the harm, the higher the verification threshold	Evidence type (primary vs hearsay); corroboration count; context integrity (full clip vs edited); certainty language without proof

Stage 3 Transmit or Respond (Speech integrity and harm prevention)	Q 33/70; Q 49/11–12; Q 24/11–19; Q 5/8; Q 49/10	Preserve dignity, justice, and cohesion in public discourse	Share only verified content; correct misinformation with adab (أدب) and without humiliation; avoid mockery and identity contempt; prefer lawful procedures in reputational matters; adopt de-escalation and reconciliation norms	Reputational violence, doxxing, public shaming; hate speech through mockery and collective blame; incitement and escalation language	If verification is incomplete or harm is likely, the ethically correct response is restraint if correction is needed, it must be truthful, dignified, and non-inciting	Dignity violations (mockery, name-calling); collective blame; incitement cues; correction quality (civil vs humiliating); retraction behavior
Cross-cutting ethical constraints (Apply at all stages)	Q 49/11–12; Q 5/8; Q 24/15–16	Ensure justice and equal standards regardless of group identity	Apply the same evidentiary threshold to in-group and out-group claims; reject suspicion-driven interpretations; avoid	Sectarian polarization; partisan double standards; outrage incentives	Justice and verification standards must not change with identity or hostility	Double standards in evidence acceptance; suspicion framing; dehumanizing labels; refusal to retract when disproven

			rumor normalization			
Applied output (*Tabayyun* Protocol)	Q 49/6; Q 17/36; Q 33/70; Q 24/11–19	Convert Qurʾānic ethics into a practical digital "workflow"	Pause and classify source audit evidence and context audit harm audit respond ethically or refrain	WhatsApp rumor cascades; "religious warning" virality; accusation culture	The minimum ethical workflow is Pause, Verify, Assess Harm, Then Share or Refrain	Presence of verification steps; links to credible sources; warnings against forwarding; corrective messaging tone

In this table, the Qurʾānic concept of *tabayyun* (تبيّن) is operationalized by packaging digital ethics into three lifecycle stages of receiving, verifying, and transmitting or responding. All levels associate Qurʾānic anchors with typical patterns of risk associated with Pakistan-relevant features like virality pressures, anonymity, reputational attacks, and identity contempt. The "decision rule" column helps to convert the Qurʾān's moral thinking into judgments user-friendly enough to influence real-world everyday online conduct. The last column offers visible clues that may serve to code and analyze actual misinformation and hate-speech-border cases in a transparent qualitative approach.

2.3 Contextual integration for Pakistan and analytic operationalization:

The QDEMT is increasingly urgent in the practical context of Pakistan with the emerging digital environment. Trends in national connectivity indicate high percentages of social media penetration and high mobile-first use, both of which are linked to high diffusion rates and low resistance forwarding.[1] The systematic online harassment and identity-based abuse also have been registered by the civil-society surveillance, meaning that the phenomena of misinformation and hate speech are not

[1] Kepios, "Digital 2024 Pakistan," *DataReportal* (We Are Social and Meltwater, 2024), https://datareportal.com/reports/digital-2024-pakistan.

isolated incidents but patterns of networked interaction.[1] Based on this, QDEMT is meant to be methodologically usable: it provides a codebook logic that would be used to conduct content analysis later, that is, by connecting Qurʾānic principles with the operational indicators (source traceability, evidence quality, language of certainty, identity contempt markers, calls to mobilization, and correction quality). In this way, the model not only safeguards the purity of the Qurʾānic moral reasoning but also enables the methodical examination of modern digital artifacts, placing the study of the Qurʾān firmly within the problematic framework of communicative issues in Pakistan.

3. Literature Review:

The information pertinent to *tabayyun* (تبيّن) and modern information damages can be clustered into four overlapping strands: (i) classical Qurʾānic exegesis on verification and reputational protection, (ii) prophetic guidance and Islamic ethical literature on the role of speech and transmission, (iii) modern studies of the relationship between misinformation and hate speech online, and (iv) digital rights and the media ecology studies that apply specifically to Pakistan.

First, Q 49/6 is used as an epistemic warning text on the tradition of the tafsīr concerning the life of the community. The classical exegetes stress that the verse connects the acts of reception of news to moral responsibility since unverified *nabaʾ* may generate the misdeeds of the other person and regret afterwards. The discussion of al-Zamakhsharī is a premonition of the ethical justification of verification before consequential action, whereas the interpretation of the verse offered by al-Bayḍāwī and others is a rule-like stance of preliminary prudence where the origin and the possible damage are not well known.[2] The edifyingly informative code of rumor circumvention and the social depreciation of accusatory speech by adopting the technique of rumor recitation (*ifk* passage; Al-Qurʾān: 24/11-19) has also been explored by exegetes as a paradigmatic condemnation within the equal communicative-ethics domain, as it then provides a Qurʾānic model that can be directly transferable in relation to digital contexts of "viral allegations."[3]

---

[1] Digital Rights Foundation (DRF), *Cyber Harassment Helpline Annual Report 2023* (Islamabad: DRF, 2023), https://digitalrightsfoundation.pk/publications/.

[2] Abū al-QāsimMaḥmūd ibn ʿUmar al-Zamakhsharī, *al-Kashshāfʿanḥaqāʾiqghawāmiḍ al-tanzīl* (تفسير الزمخشري: الكشاف عن حقائق غوامض التنزيل), commentary on Q 49:6; Nāṣir al-Dīn al-Bayḍāwī, *Anwār al-tanzīlwa-asrār al-taʾwīl* (تفسير البيضاوي: أنوار التنزيل وأسرار التأويل), commentary on Q 49:6.

[3] Ibn Kathīr, *Tafsīr al-Qurʾān al-ʿAẓīm*, commentary on Q 24:11–19..

Complementary to this, Q 49/11-12 grounds a dignity-based ethic that forbids ridicule and mistrust as well as criticizing persons as a form of moral violence.

Second, the insistence on verification as practiced by the Qurʾān is reinforced by Hadith literature by viewing careless transmission as a type of untruthfulness in itself. The report that features commonly in*Ṣaḥīḥ Muslim*—"It is enough falsehood for a man to narrate everything he hears"—assures the formal pragmatic principle of prophetic pitting against indiscriminate transmitting that is akin to the morality-logic of *tabayyun* in the networked environment.[1] Similarly, *Ṣaḥīḥ al-Bukhārī* cautions against suspicion as "the most false of speech," and it correlates with Q 49/12 and indicates that interpretive practices due to suspicion may be morally flawed even prior to the expression of hate or accusation.[2] In addition to hadith, classical sources of ethics, including the *Iḥyāʾʿulūm al-dīn* by al-Ghazālī, can provide a rich moral psychology of the tongue—how social desirability, anger, or piety-performance may pervert speech into *ghībah* (غيبة) and *namīmah* (نميمة)—helpful in understanding the attractivity of misinformation and hate under polarized circumstances.[3]

Third, modern research on communication and computational social science describes the diffusion of misinformation and moralized hostility in the digital setting driven by digital incentives. Empirical research can prove that "fake news" can spread faster than true news through online networks, partly because of novelty and an emotional reaction.[4] Discussions of scholarship about fake news and political misinformation focus more on the structural position of platform affordances and attention economies than on a situation of ignorance among individuals.[5] Evidence on moral-emotional language suggests that outrage and moral condemnation make diffusion rise even more—evidence that sheds light on why religiously encrypted

[1] Muslim ibn al-Ḥajjāj, *Ṣaḥīḥ Muslim*, introduction, hadith no. 5, accessed February 18, 2026, https://sunnah.com/muslim:5.

[2] Muḥammad ibn Ismāʿīl al-Bukhārī, *Ṣaḥīḥ al-Bukhārī*, "Good Manners and Form," hadith no. 6064, accessed February 18, 2026, https://sunnah.com/bukhari:6064.

[3] AbūḤāmid Muḥammad al-Ghazālī, *Iḥyāʾʿulūm al-dīn* (إحياء علوم الدين), Book on "Āfāt al-lisān" (آفات اللسان). For an accessible Arabic text, see https://archive.org/details/IhyaUlumAdDinGhazzali.

[4] Soroush Vosoughi, Deb Roy, and Sinan Aral, "The Spread of True and False News Online," *Science* 359, no. 6380 (2018): 1146–51, https://doi.org/10.1126/science.aap9559.

[5] Hunt Allcott and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives* 31, no. 2 (2017): 211–36, https://doi.org/10.1257/jep.31.2.211.

content about outrage is frequently very shareable.[1] These literatures assist in defining the "risk environment" in which Qurʾānic standards are applicable.

Finally, the local online harm ecology is recorded through Pakistan-based reports. In Freedom House, limitations, harassment, and manipulation of information in the Pakistani digital environment are noted, and the social costs of being abused and being in hostilities related to identity online are reported by civil society research.[2] Nevertheless, the difference between descriptive digital-rights reporting and Qurʾān-based analytical models that will inform Muslim ethical agency in the information lifecycle persists. This article argues that *tabayyun*, in combination with the Qurʾānic ethics of dignity and justice, is a stringent conceptual intermediate between the study of the Qurʾān and practical contemporary problems of communication.

4. Methodology, Data Collection, and Analysis:

This paper uses a qualitative, two-layered research design synthesizing (i) thematic Qurʾānic analysis (*tafsīrmawḍūʿī*) and (ii) qualitative content analysis of digitally oriented artifacts of Pakistan. The first layer constitutes the development of the QDEMT framework, the thematic grouping of passages in the Qurʾān on verification, accountable knowledge, disciplined speech, and the protection of dignity, justice, and reconciliation (e.g., Q 49/6; Q 17/36; Q 33/70; Q 49/11-12; Q 24/11-19; Q 5/8; Q 49/10). Thematic Qurʾānic analysis here is operationalized as follows: (a) the cluster of verses by ethical concept, (b) the concentration of semantics by maximally significant terms (e.g., *tabayyun*, *nabaʾ*, *qawl*, *buhtān*, *ʿadl*), and (c) triangulation with representative classical tafsīr discourse to prevent the problem of decontextual reading.[3]

In the second layer, the study creates a small, auditable sample of publicly accessible instances of misinformation and hate-speech-adjacency material concerning Pakistan, which is purposely chosen by sampling cases already documented and stored in reputable fact-checking and monitoring organizations (to ensure traceability and

[1] William J. Brady et al., "Emotion Shapes the Diffusion of Moralized Content in Social Networks," *Proceedings of the National Academy of Sciences* 114, no. 28 (2017): 7313–18, https://doi.org/10.1073/pnas.1618923114.

[2] Freedom House, "Pakistan Freedom on the Net 2023," accessed February 18, 2026, https://freedomhouse.org/country/pakistan/freedom-net/2023.

[3] For a contemporary methodological discussion relevant to thematic and context-sensitive Qur'anic interpretation, see Abdullah Saeed, *Interpreting the Qur'an Towards a Contemporary Approach* (London: Routledge, 2006), p. 45.

avoid selection bias among researchers).[1] Samples include items across social platforms, news claims, accusations, mobilization posts, and identity-targeting stories.

This is analyzed through a systematic procedure of qualitative content analysis: (i) developing a codebook based on QDEMT pillars; (ii) coding all items in terms of source-evidence quality, certainty language, or harm probability, contempt indicators of dignity, and violence of identity; (iii) mapping the patterns of coding in relation to Qurʾānic breaches of ethics and probable social harms.[2] Transparent coding rules and the development of an audit trail of sources increase reliability. Method decisions are based on the guidelines of content analysis methodology and research integrity principles regarding sensitive material.[3]

Table 2 Analytic Codebook for Qualitative Content Analysis Using QDEMT

Code	What is coded (definition)	Qurʾānic grounding (concept and anchors)	Inclusion indicators (what counts)	Notes for consistent coding
C1 Source opacity	Origin of claim cannot be traced to a credible, identifiable source	*Tabayyun* Q 49/6; *ʿilm* Q 17/36	Anonymous screenshots; "someone said"; forwarded messages with no origin	Code even if the claim later proves true the issue is traceability at time of sharing
C2 Evidence absence or weakness	Claim lacks verifiable proof or uses misleading	Q 17/36; Q 49/6	No primary documents; out-of-context photos; edited clips;	Record evidence type photo, clip, testimony,

[1] For transparency and traceability in contested claims, the study relies on documented case archives such as: Soch Fact Check, "Soch Fact Check," accessed February 22, 2026, https://www.sochfactcheck.com/; Agence France-Presse (AFP), "AFP Fact Check Pakistan," accessed February 22, 2026, https://factcheck.afp.com/afp-pakistan.

[2] Klaus Krippendorff, *Content Analysis An Introduction to Its Methodology*, 4th ed. (Thousand Oaks, CA: SAGE Publications, 2019), https://us.sagepub.com/en-us/nam/content-analysis/book258450; Margrit Schreier, *Qualitative Content Analysis in Practice* (London: SAGE Publications, 2012), https://methods.sagepub.com/book/mono/qualitative-content-analysis-in-practice/toc# .

[3] Committee on Publication Ethics (COPE), "Core Practices," accessed February 26, 2026, https://publicationethics.org/core-practices.

	"proof"		uncited statistics	official statement, etc.
C3 Certainty without proof	High-confidence language presented without evidence	*Qawlsadīd* Q 33/70; Q 24/15–16	"Confirmed," "exposed," "100% true" without citations	Distinguish between "certainty" and "opinion" phrasing
C4 Religious coercion framing	Claim pushes sharing as a religious test or duty	Q 33/70; Q 24/15–16	"Share if you love Islam"; "don't question this"; takfīr-like insinuations	Code regardless of topic (political, sectarian, medical) if coercive piety is used
C5 Accusatory rumor	Allegation against a person or group circulated without due process	*Ifk* ethics Q 24/11–19; dignity Q 49/11–12; evidentiary seriousness Q 24/4	Name-and-shame posts; moral accusations; implied guilt through insinuation	Treat as "high-harm" content note if private individuals are targeted
C6 Dignity and privacy violation	Content exposes private information or humiliates targets	Dignity *'irḍ* Q 49/11–12	Doxxing; intimate images; mocking captions; degrading memes	Code alongside C5 when accusations also exist
C7 Mockery and name-calling	Derision or stigmatizing labels used against identities	Q 49/11	Slurs; insulting nicknames; ridicule markers; "disgusting people"	Note identity target type sect, ethnicity, gender, political group

C8 Collective blame and dehumanization	A whole group is blamed or described as less than human	Justice Q 5/8; dignity Q 49/11	"They are all…"; animal metaphors; calls to expel/ban collectively	Code even when framed as "critique" if it generalizes and dehumanizes
C9 Incitement and mobilization cues	Content encourages hostility, retaliation, or unlawful action	Cohesion *iṣlāḥ* Q 49/10; justice Q 5/8	"Teach them a lesson"; calls for violence; mob triggering; targeted harassment	Note whether mobilization is immediate ("go now") or general ("they deserve…")
C10 Selective verification	Evidentiary standards change with group identity or partisan loyalty	Justice Q 5/8; Q 17/36	Weak evidence accepted for out-group; strong evidence demanded for in-group	Code via comparative cues within the same thread/channel when visible
C11 Ethical correction with adab	Corrections are offered truthfully without humiliation	*Qawlsadīd* Q 33/70; dignity Q 49/11–12	Calm correction; links to sources; respectful language	Mark as "protective factor" when present
C12 Retraction responsibility	Whether original sharer/page retracts or clarifies after correction	Accountability Q 17/36; Q 24/15–16	Public correction, deletion with note, refusal to retract	Useful for distinguishing harmful persistence from good-faith error

This codebook breaks down QDEMT into a useful collection of qualitative codes used for posts, screenshots, clips, captions, and comment threads about Pakistan. All codes are based on a Qurʾānic ethical category: verification, accountable knowledge, disciplined speech, dignity protection, justice, and reconciliation, due to

which the analysis will be Qurʾān-centered, not purely sociological. The "indicators of inclusion" offer a stable set of criteria to code by, which allows you to be transparent and have an audit trail in your method. The last two codes (ethical correction and retraction) are necessary so that the analysis captures not only harms but also Qurʾān-oriented corrective actions that minimize misinformation and hate.

5. Discussion and Analysis:

It uses the Qurʾānic Digital Ethics Model of *Tabayyun* (QDEMT) to structure a purposively selected collection of Pakistan-relevant fake news and hate-speech-related content that was publicly stored by mainstream fact-checking and monitoring projects. The analyses are not meant to provide estimates of prevalence; instead, they present recurring patterns of communication and explain them in terms of Qurʾānic categories of ethics that cover (i) reception, (ii) verification, and (iii) transmission/response. The central argument is that the "high-velocity" digital space in Pakistan creates the conditions in which the Qurʾānic necessity of *tabayyun* (تبيّن) becomes a matter of moral obligation and a systemically operating.

5.1 Pattern One: Virality replacing verification at the reception stage:

The most frequent one is speed: the posts are marked as urgent ("forward to all," "before it gets deleted"), and users are framed as moral actors by immediately sharing. Contemporary research describes the bias of the online attention economy to prefer novelty and emotion in such a way that inaccurate assertions spread more than their attempts at correction.[1] As empirical research shows, false news can spread more quickly than the truth in networked environments, as it is usually more recent and emotionally provoking.[2] These dynamics once again collide squarely with Qurʾānic epistemic limitation: Q 49/6 connects the context of receiving *nabaʾ* (news) as the application of the moral duty of clarification to prevent unwarranted harm, whereas Q 17/36 defines the context of accepting unfounded claims as morally liable. Classical tafsīr focuses on the fact that the interpretation of verses is not to engage in unjust action based on incoherent data—a moral logic that can easily be traced to any contemporary "share-first" harms.[3]

[1] Cass R. Sunstein, *Republic Divided Democracy in the Age of Social Media* (Princeton, NJ: Princeton University Press, 2017), https://press.princeton.edu/books/hardcover/9780691175515/republic.
[2] Soroush Vosoughi, Deb Roy, and Sinan Aral, "The Spread of True and False News Online," *Science* 359, no. 6380 (2018): 1146–51, https://doi.org/10.1126/science.aap9559.
[3] al-Qurṭubī, *al-Jāmiʿ li-aḥkām al-Qurʾān*, commentary on Q 49:6.

QDEMT reflects this tendency in disaggregating the ethics of reception, in which exposure is quickly converted to judgment and action. Qurʾānic counteraction is not a response to skepticism but moral prudence, that is, to categorize the affirmation (news, accusation, mobilization), to find predictively harmful results, and to trace the origin of the affirmation as a sign to evade it.

5.2 Pattern Two: Piety-coded misinformation and moral-emotional diffusion:

The second inclination is the dissemination of falsehoods under the guise of religious exhortation—"Islamic warning," "proof of fitnah," "defend honor"—which justifies opinions against criticism by putting verification as treachery. Empirical results of moralized content show that moral-emotional language use (with outrage in particular) enhances the spread within social networks.[1] This is one of the reasons why religiously coded outrage can be so infectious even in cases of weak evidence.

In Qurʾānic ethics, piety, however, is never addressed outside of truthfulness: *qawlsadīd* (قول سديد) in Q 33/70 requires restrained accuracy and not performative certainty. Similarly, Q 24/15-16 also criticizes rumor-circulation involving the weighty allegation as "a trivial matter," revealing a moral psychology of social talk that is morally insignificant. In the classical exegetical tradition, this passage can be understood as a community reproach: the readers who transmit it are in the business of the moral lapse, although not the initiators of the assertion.[2] At the level of operational intervention, QDEMT coded piety-coded misinformation using such markers as the language of religious coercion, evidence-free certainty, and the appeal to mass-forward as an act of faith. Religious commitment needs to shift its emotional expression into accountable verification, which is the Qurʾānic correction.

5.3 Pattern Three: Reputational violence and the Qurʾānic dignity paradigm:

The third and socially threatening trend is reputational violence: accusations are made against someone or a group of people through screenshots and edited videos or anonymous posts, and then a group is humiliated and protested against publicly. This pattern is particularly harmful due to the context of Pakistan, where cases of moral accusations may lead to a faster spiral. The dignity-defending form of the

[1] William J. Brady et al., "Emotion Shapes the Diffusion of Moralized Content in Social Networks," *Proceedings of the National Academy of Sciences* 114, no. 28 (2017): 7313–18, https://doi.org/10.1073/pnas.1618923114.

[2] Ibn Kathīr, Tafsīr al-Qurʾān al-ʿAẓīm, commentary on Q 24:15–16.

Qurʾān is clear: Q 49/11-12 outlaws mockery, name-calling, suspicion (*ẓann*), and backbiting (*ghībah*), whereas Q 24/4 sets the evidentiary seriousness of accusations very high. The *ifk* story (Al-Qurʾān: 24/11-19) serves as a Qurʾānic "communication ethics case": not only the failure of fabrication but also communication (transmission), perfection (normatizing), and a denial of principled restraint.[1]

In QDEMT, reputational claims are accorded the status of "high-risk material" that requires intensified *tabayyun* and, most importantly, procedural restraint: in those circumstances, when the allegations concern dignity and social order, the ethical reaction is not a viral rotation but the legal procedure and cautious correction. This is in line with the Prophetic guidance that good speech is a state of faith—"Whoever believes in Allah and the Last Day should speak good or remain silent."[2] Therefore, the model describes the comment culture and quote-post punishment as not a neutral practice but as a potentially unjustified intervention in the *buhtān* (بهتان) and infringement of dignity.

5.4 Pattern Four: Hate speech as contempt, collective blame, and justice failure:

Identity-targeting texts are also present in the sampled corpus, no longer aimed at ideas, but at people—teasing, dehumanizing names, withering the crowd, hints at betrayal. UN guidance identifies hate speech as the promotion of hostility, which probably will lead to discrimination and violence, particularly when group resentment is being organized.[3] This kind of language can be a social bomb; polarization and sectarianism can help it become a bomb in Pakistan.

The Qurʾān addresses ethics at various levels. Q 49/11 denies ridicule and derogatory classification (*tanābuzbil-alqāb*), defining the dignity floor for public speech. Q 5/8 thereafter contains a justice ceiling: aggression against a people should not lead to injustice. Practically, hate speech tends to be closely paralleled by selective verification—giving in to dusty statements when they are confirmatory. Contemporary misinformation research has proposed that the presence alone of basic "accuracy prompts" can minimise the spread of false accusations, and that it is the lack of attention to accuracy, which is often a behavioural rather than an informational

[1] al-Ṭabarī, Jāmiʿ al-bayān, commentary on Q 24:11–19.
[2] Muḥammad ibn Ismāʿīl al-Bukhārī, *Ṣaḥīḥ al-Bukhārī*, "Good Manners and Form," hadith no. 6018, accessed February 28, 2026, https://sunnah.com/bukhari:6018.
[3] United Nations. *United Nations Strategy and Plan of Action on Hate Speech* (New York: United Nations, 2019).

phenomenon.[1] QDEMT considers selective verification as an ethical failure of *ʿadl* (عدل) and epistemic integrity: the same amount of evidence must be applied regardless of one's or the other group's identity. Operational codes are dehumanizing metaphors, language of collective blame, and a double standard in evidentiary demand.

5.5 The *Tabayyun* Protocol as an applied output of Qurʾānic studies:

The main addition of this article is that the Qurʾānic morality words are redefined into a more realistic, responsible process within the context of online behavior: (i) Pause and classify the statement (news, accusation, mobilization, religious warning), (ii) Source audit (plan of origin, trace, credibility, conflict of interest), (iii) Evidence/context audit (primary evidence, support, and media edits), (iv) Harm audit (dignity, sectarian tension, public safety, possible incitement) and (v) Response ethics (refrain if unverified; correct with *adab* (أدب); avoid humiliation; prefer lawful channels for accusations).

This protocol conforms to more general "infodemic" municipal remedies that aim to slow transmission, limit verification, and reduce harm during crises, but anchors these solutions in Qurʾānic taxonomies rather than rigorously technocratic logic.[2] Communication, in Qurʾānic terms, is the subject of the protocol as *amānah* (trust): a person is not only responsible but also involved in the process of diffusion. The Qurʾānic principle of validation and dignity may be used against digital surveillance capitalism, in which outrage is the incentive and in which attention-setting works as counter-logic,[3] by reframing justice, moderation, and reconciliation (*iṣlāḥ*, إصلاح) as communicative objectives in Pakistan's digital world.

6. Conclusion:

This paper concludes that the Qurʾānic concept of *tabayyun* (تبيّن) in Q 49/6 develops a strong epistemic ethic of digital life by asserting that verification is also a moral obligation whenever the reception or transmission of nabaʾ (news) can cause harm. When the Qurʾān is read in conjunction with Q 17:36, the online engagements which include reading, reacting, commenting, and forwarding, are placed in a morally responsible framework instead of consumer neutrality. It is also demonstrated in the

[1] Gordon Pennycook et al., "Shifting Attention to Accuracy Can Reduce Misinformation Online," *Nature* 592 (2021): 590–95, https://doi.org/10.1038/s41586-021-03344-2.

[2] World Health Organization, "Infodemic," accessed March 4, 2026, https://www.who.int/health-topics/infodemic.

[3] Shoshana Zuboff, *The Age of Surveillance Capitalism The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, 2019), 66.

analysis that the Qurʾānic ethic of disciplined speech (*qawlsadīd*, Q 33/70) and the ifk template (Al-Qurʾān: 24/11-19) specifically deal with the mechanics of viral rumor, by relying on collective responsibility to circulate and rectify rumors. Lastly, the norms of dignity and justice outlined in the Qurʾān(Al-Qurʾān: 49/11-12; Q 5/8) offer an explicit set of ethical parameters against mockery, identity loathing, collective censure, and discriminatory verification, and thus can serve as a principled approach to address the issue of hate speech and polarization of digital space in Pakistan.

The article concludes that Qurʾān-based digital ethics does not necessarily rely on moral exhortation; it can be implemented by adopting the Qurʾānic Digital Ethics Model of *Tabayyun* (QDEMT), which governs the entire information lifecycle—receive, verify, transmit, or respond. QDEMT provides a bridging, analytically elliptical statement between Qurʾān studies and modern social troubles in the context of Pakistan, where virality, anonymity, and identity-based contestation can increase harm to communicative functions and judgments by turning Qurʾānic norms into practices and evaluative measurements.

This paper suggests adhering to a license to check the authenticity of the post before reposting; adhering to adumbration, not to defame, but instead to receive the correction with adab (db); and adhering to defamation and corrections with *adab* (أدب). It recommends combining the Q 49 and Q 24 communication ethics into khutbah discourse, discourages the circulation of rumors in the community as a community sin, and encourages the ethics of justice in disagreeing. It also suggests introducing madaris and university programs (along with the process of their practical transfer), the element of "digital *adab* literacy", and the principles of online speech ethics. It also recommends carrying out the establishment of fact-checking alliances and community media-literacy efforts and providing free and equal reaction to hate speech and disinformation concerning dignity and societal integrity, without providing a platform to misuse this authority.

7. Suggestions for Future Research:

There are four ways in which future research can build on this study. First, to address the research question, awareness of *tabayyun* (تبيّن) in relation to everyday verification behavior and predictors of unverified forwarding as the dependent variable can be measured in a quantitative survey across the provinces of Pakistan and across age groups, then correlated with religiosity, education, and media-consumption patterns. Second, scholars may formulate and authenticate an empirical QDEMT scale to evaluate that shared phenomenon, even as a measure of "Qurʾān-aligned digital

ethics," and test the hypothesis of whether media-literacy interventions are effective in obtaining better results over time. Third, comparative analysis could be used to understand how the workings of *tabayyun*-centred digital ethics can vary within Muslim-majority societies with dissimilar platform ecologies and sectarian structures, as well as the meaning of what is culturally particular and generalizable. Fourth, interdisciplinary efforts would then incorporate categories of Qurʾānic ethics into computation-based identification of Urdu and Roman-Urdu hate speech and rumor clues, and at the same time, analyzing the norms of correction, retraction, and reconciliation-focused communication practices in online social networks.

Bibliography:

Agence France-Presse (AFP). *AFP Fact Check Pakistan*. Accessed February 22, 2026. https://factcheck.afp.com/afp-pakistan.

Allcott, Hunt, and Matthew Gentzkow. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31, no. 2 (2017): 211–236. https://doi.org/10.1257/jep.31.2.211.

al-Bayḍāwī, Nāṣir al-Dīn. *Anwār al-Tanzīlwa-Asrār al-Taʾwīl (تفسير البيضاوي)*.

al-Bukhārī, Muḥammad ibn Ismāʿīl. *Ṣaḥīḥ al-Bukhārī*.

al-Ghazālī, AbūḤāmid Muḥammad. *Iḥyāʾ ʿUlūm al-Dīn (إحياء علوم الدين)*.

al-Qurṭubī, AbūʿAbd Allāh Muḥammad ibn Aḥmad. *Al-Jāmiʿ li-Aḥkām al-Qurʾān (الجامع لأحكام القرآن)*.

al-Rāzī, Fakhr al-Dīn. *Mafātīḥ al-Ghayb (مفاتيح الغيب)*.

al-Ṭabarī, Muḥammad ibn Jarīr. *Jāmiʿ al-BayānʿanTaʾwīlĀy al-Qurʾān (جامع البيان عن تأويل آي القرآن)*.

al-Zamakhsharī, Abū al-QāsimMaḥmūd ibn ʿUmar. *Al-KashshāfʿanḤaqāʾiqGhawāmiḍ al-Tanzīl (الكشاف عن حقائق غوامض التنزيل)*.

Brady, William J., Molly J. Crockett, Jay J. Van Bavel, and others. "Emotion Shapes the Diffusion of Moralized Content in Social Networks." *Proceedings of the National Academy of Sciences* 114, no. 28 (2017): 7313–7318. https://doi.org/10.1073/pnas.1618923114.

Committee on Publication Ethics (COPE). *Core Practices*. Accessed February 26, 2026. https://publicationethics.org/core-practices.

Digital Rights Foundation. *Cyber Harassment Helpline Annual Report 2023*. Islamabad: Digital Rights Foundation, 2023. https://digitalrightsfoundation.pk/publications/.

Freedom House. *Freedom on the Net 2023: Pakistan*. Accessed February 18, 2026. https://freedomhouse.org/country/pakistan/freedom-net/2023.

Ibn Kathīr, Ismāʿīl ibn ʿUmar. *Tafsīr al-Qurʾān al-ʿAẓīm (تفسير القرآن العظيم)*.

Ireton, Cherilyn, and Julie Posetti, eds. *Journalism, "Fake News" & Disinformation: A Handbook for Journalism Education and Training*. Paris: UNESCO, 2018. https://unesdoc.unesco.org/ark:/48223/pf0000265552.

Kepios. *Digital 2024: Pakistan*. DataReportal. We Are Social and Meltwater, 2024. https://datareportal.com/reports/digital-2024-pakistan.

Krippendorff, Klaus. *Content Analysis: An Introduction to Its Methodology*. 4th ed. Thousand Oaks, CA: SAGE Publications, 2019.

Muslim ibn al-Ḥajjāj. *Ṣaḥīḥ Muslim*.

Office of the United Nations High Commissioner for Human Rights (OHCHR). *Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred That Constitutes Incitement to Discrimination, Hostility or Violence*. Geneva: OHCHR, 2012. https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf.

Pennycook, Gordon, David G. Rand, and others. "Shifting Attention to Accuracy Can Reduce Misinformation Online." *Nature* 592 (2021): 590–595. https://doi.org/10.1038/s41586-021-03344-2.

Saeed, Abdullah. *Interpreting the Qur'an: Towards a Contemporary Approach*. London: Routledge, 2006.

Schreier, Margrit. *Qualitative Content Analysis in Practice*. London: SAGE Publications, 2012.

Soch Fact Check. *Soch Fact Check*. Accessed February 22, 2026. https://www.sochfactcheck.com/.

Sunstein, Cass R. *Republic Divided: Democracy in the Age of Social Media*. Princeton, NJ: Princeton University Press, 2017.

United Nations. *United Nations Strategy and Plan of Action on Hate Speech*. New York: United Nations, 2019.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. "The Spread of True and False News Online." *Science* 359, no. 6380 (2018): 1146–1151. https://doi.org/10.1126/science.aap9559.

Wardle, Claire, and Hossein Derakhshan. *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*. Strasbourg: Council of Europe, 2017.

World Health Organization. *Infodemic*. Accessed March 4, 2026. https://www.who.int/health-topics/infodemic.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs, 2019.